10 Fedorov, A. *et al.* (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16128–16133
11 Rogozin, I.B. *et al.* (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13, 1512–1517
12 Roy, S.W. and Gilbert, W. (2005) Complex early genes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1986–1991
13 Koonin, E.V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct* 1, 22 DOI: 10.1186/1745-6150-1-22 (www.biology-direct.com)
14 Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus–cytosol compartmentalization. *Nature* 440, 41–45
15 Doolittle, W.F. (1978) Genes in pieces: were they ever together? *Nature* 272, 581–582
16 Gilbert, W. and Glynias, M. (1993) On the ancient nature of introns. *Gene* 135, 137–144
17 Rogers, J.H. (1990) The role of introns in evolution. *FEBS Lett.* 268, 339–343
18 Lambowitz, A.M. and Zimmerly, S. (2004) Mobile group II introns. *Annu. Rev. Genet.* 38, 1–35
19 Robart, A.R. and Zimmerly, S. (2005) Group II intron retroelements: function and diversity. *Cytogenet. Genome Res.* 110, 589–597
20 Makarova, K.S. *et al.* (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* 33, 4626–4638
21 Sverdlov, A.V. *et al.* (2005) Conservation versus parallel gains in intron evolution. *Nucleic Acids Res.* 33, 1741–1748
22 Babenko, V.N. *et al.* (2004) Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Res.* 32, 3724–3733

0168-9525/$ – see front matter
doi:10.1016/j.tig.2007.01.001

Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor

Marie Sémon and Kenneth H. Wolfe

Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland

The whole genome duplication that occurred in ray-finned fish coincided with the radiation of teleost species; consequently, these two phenomena have often been linked. Using the *Tetraodon* and zebrafish complete genome sequences, we tested a molecular hypothesis that can relate whole genome duplication to speciation in teleosts. We estimate that thousands of genes that remained duplicated when *Tetraodon* and zebrafish diverged underwent reciprocal loss subsequently in these two species, probably contributing to reproductive isolation between them.

Introduction

There is now conclusive evidence that a whole genome duplication (WGD; see Glossary) occurred in ray-finned fish [1–4]. Because the timing of this WGD and the radiation of teleost species approximately coincided, it has often been suggested that these two phenomena are associated [2,3,5–7]. Here, we test a molecular hypothesis that could link WGD to speciation in teleosts.

After WGD, many genes are deleted so that the gene content is only slightly increased [1]. Previous analysis in yeast species showed that reciprocal gene loss (RGL) occurs at duplicated loci and contributes to the speciation process [8]. RGL occurs when two paralogs created by WGD are retained until speciation, after which each species loses a different copy. In diploid species, such as teleosts, each pair of paralogs that undergoes RGL in two lineages results in one-sixteenth of the F2 zygotes of an interlineage mating being nonviable (if the remaining paralog is essential). Therefore, the neutral loss of copies of duplicated genes is a powerful lineage-splitting force (as a result of Dobzhansky–Muller incompatibilities [9–12]). To evaluate whether losses of alternative copies of duplicated genes contributed to the evolution and speciation of teleosts, we estimated the extent of RGL between *Tetraodon nigroviridis* (called *Tetraodon* from this point) and zebrafish (*Danio rerio*), the complete genome sequences of which are available in Ensembl [1,13]. Although there is considerable uncertainty about the precise timing of both the WGD {estimated from molecular data to be 253–404 million years ago (Mya) [3,5]} and the divergence between the lineages leading to *Tetraodon* and zebrafish {estimated from fossil data to be 150–165 Mya [14] and from molecular data to be 254–324 Mya [15–17]}, these species are the most appropriate for this analysis because they have the most divergent pair of teleost genomes that have been sequenced and because the time since their divergence accounts for a significant proportion of the time since WGD.

A two-to-one mapping relationship between teleost and outgroup genomes

We define an ancestral locus as a locus that existed in the last common ancestor of tetrapods and ray-finned fish. A set of modern-day loci is descended from this ancestral locus. This set of descendants comprises the orthologous locus in any outgroup species that did not undergo WGD (such as humans and chickens) and the pairs of corresponding co-orthologous loci in species that underwent WGD (such as *Tetraodon* and zebrafish). In Figure I in Box 1, each column represents the descendants of an ancestral locus. At some ancestral loci, the descendants in the species

Corresponding author: Wolfe, K.H. (khwolfe@tcd.ie).
Available online 1 February 2007.

Glossary

Ancestral locus: a locus that existed in the most recent common ancestor of the genomes being compared and that has descendant loci in these genomes. An ancestral locus has one descendant locus in any species that did not undergo WGD and two descendant loci in any species that underwent WGD. However, intact genes might not be retained at all the descendant loci.

Divergent resolution (differential gene loss): resolution refers to the long-term outcome of a gene duplication event, resulting in the survival of two, one or (rarely) no copies of the gene in a species. Divergent resolution means that two species show different outcomes from a gene duplication event that occurred in their common ancestor: for example, one species retains both gene copies, and the other species retains only one gene copy. RGL is a particular example of divergent resolution. The terms divergent resolution and differential gene loss are synonymous.

Dobzhansky–Muller incompatibilities: Dobzhansky, Muller and Bateson proposed that incompatible interactions between alleles segregating at two loci in different populations can reduce the fitness of a hybrid. In this case, if a pair of duplicated genes has undergone RGL in two lineages, an F1 hybrid would be heterozygous for a null (gene deletion) allele at each of the two loci. One-sixteenth of the F2 hybrids formed by random mating among the F1 population would be homozygous null at both loci, which would be lethal if the gene product is essential.

Interleaved: a characteristic feature of WGD is that it creates an interleaved relationship between the gene order in an outgroup species and the gene orders in the corresponding pair of genomic regions in a species that underwent WGD. For example, if the gene order in the outgroup (e.g. humans) is A-B-C-D-E, the gene orders in the two daughter regions in a post-WGD fish species might be A-C-D and A-B-E. Therefore, gene A has been retained in duplicate, and the others are single copy. In this way, the gene order in humans can be described as an interleaving of orthologs of genes on the two fish chromosomes.

Non-RGL locus: in this study, an ancestral locus that corresponds to single-copy genes in both *Tetraodon* and zebrafish and for which these single-copy genes can be confidently inferred to be orthologs. RGL has not occurred at this locus. Loss of one gene copy might have occurred in the common ancestor of the two fish after WGD, or the orthologs might have been lost independently in each fish after speciation.

Orthologs: two genes (or DNA regions) that diverged from a common ancestral gene (or region) in a speciation event so that the history of the gene reflects the history of the species.

Orthologous chromosomes: a pair of chromosomes, the common ancestor of which dates to a speciation event. For many zebrafish chromosomes, there is an orthologous *Tetraodon* chromosome (Table S2 in the supplementary material online). For other chromosomes, this simple relationship has been partly scrambled by interchromosomal rearrangements that occurred in one of the species after they diverged.

Paralogs: two genes (or DNA regions) that diverged from a common ancestral gene (or region) in a duplication event.

Paralogons: large chromosomal regions that were formed by a single duplication event. They usually contain numerous paralogous genes.

Paralogous chromosomes: a pair of chromosomes, the common ancestor of which dates to a WGD event rather than a speciation event.

Reciprocal gene loss (RGL): the situation when two lineages that have inherited a gene duplication independently lose alternative members of the duplicated pair after speciation. Hence, the surviving single-copy genes are paralogs.

RGL locus: in this study, an ancestral locus, the descendants of which can be confidently inferred to have undergone RGL in *Tetraodon* and zebrafish.

Subfunctionalization: the complementary loss of the subfunctions of an ancestral gene in each of the descendant duplicate copies.

Whole genome duplication (WGD): the simultaneous acquisition of extra copies of all of the nuclear chromosomes of an organism. Recent and ancient WGD events are also called polyploidy and paleopolyploidy, respectively. WGD can arise either by autopolyploidization (doubling of a single genome) or allopolyploidization (merging of two similar genomes).

that underwent WGD are gene pairs that have been retained in duplicate. However, for most ancestral loci, only one copy of the gene exists after WGD, and the other copy has been lost. For ancestral loci that are now represented by single-copy genes in *Tetraodon* and zebrafish, we developed a method to determine whether these single-copy survivors in the two fish species are orthologs or paralogs of each other. Data sets were obtained by phylogenetic analysis and are described in Table S1 in the supplementary material online.

Initially, we mapped genes in local genomic regions in an outgroup to the corresponding regions in the teleost genomes. In teleost species, two ancient paralogous regions (indicated as track 1 and track 2 in Figure I in Box 1) typically retain few genes in common. But, compared with an outgroup such as humans, these species show a relationship by which each of the two regions in the fish contains a subset of the genes in the corresponding region of the outgroup genome (two-to-one mapping). In the outgroup genome, homologs of genes from the two fish regions are found in an interleaved manner. By contrast, we expect to see a simple one-to-one mapping when orthologous regions from the two teleosts are compared: i.e. any genomic region in zebrafish should have a single orthologous region in *Tetraodon* and vice versa.

Numerous interchromosomal rearrangements have occurred in the human lineage and, to a lesser extent, the chicken lineage [18]. These rearrangements tend to obscure the syntenic relationship between a region in an outgroup and both of the expected sister regions in teleosts. Moreover, although interchromosomal rearrangements have been relatively rare in teleost genomes, gene order has been extensively scrambled by inversions [1,19], complicating local comparisons. Therefore, in contrast to previous studies in yeast [8,20], we needed to use a less strict conservation of gene order to identify paralogous regions. By considering syntenic regions (which contain genes that reside on the same chromosome) without a strict requirement for conservation of local gene order, we created local maps with the expected one-to-two WGD signature between a genomic region in an outgroup and a pair of genomic regions in each teleost (Box 1).

Identifying RGL loci throughout the teleost genomes

We used the neighborhood approach described in Box 1 to search for ancestral loci, the descendants of which underwent RGL in *Tetraodon* and zebrafish, using both humans and chickens as outgroups representing the ancestral state. We classified an ancestral locus as an RGL locus if it fit two criteria: first, if *Tetraodon* and zebrafish each have only one gene from the original pair formed from this ancestral locus; and, second, if the surviving post-WGD single-copy genes occur on paralogous chromosomes (i.e. on different tracks). Conversely, we classified an ancestral locus as a non-RGL locus if its descendants are single copy in both *Tetraodon* and zebrafish, and if the remaining copies occur on orthologous chromosomes (i.e. on the same track). In this way, we classified about one-third of the studied ancestral loci with confidence (Table 1). We could not infer the status of an ancestral locus when the gene order in the region was too scrambled to define homologous chromosomes between outgroup and teleost species, or when a significant orthology relationship between *Tetraodon* and zebrafish chromosomes could not be determined. Combining the results obtained using human and chicken outgroups, we found that the descendants of 8% of the studied ancestral loci underwent RGL in *Tetraodon* and zebrafish (Table 1, Outgroup merged row). The 73 identified RGL loci are listed in Table S3 in the supplementary material online. A gene ontology analysis [21] of this group (using FatiGO; http://fatigo.bioinfo.cipf.es) did not reveal

Box 1. Identification of RGL at the matrilin 3 ancestral locus

We illustrate our approach to finding RGL loci in teleosts by considering the example of the fish orthologs of the human gene matrilin 3 (*MATN3*). Briefly, we identify syntenic regions (chromosomes) in *Tetraodon* and zebrafish that are putatively co-orthologous to the human region around *MATN3*, and then we seek to establish the orthology relationships between *Tetraodon* and zebrafish for the identified chromosomes.

We begin by comparing human to *Tetraodon*. We consider a neighborhood of 80 human genes that have orthologs in *Tetraodon* (40 genes upstream and 40 genes downstream of *MATN3* but excluding *MATN3* itself). It is not possible to assess this neighborhood for genes located near the ends of contigs or chromosomes, so we analyzed only loci for which at least 50 neighboring genes, including both upstream and downstream genes, were available.

In the ideal case, the orthologs of these 80 human genes should be present on exactly two *Tetraodon* chromosomes, corresponding to the two paralogons (see Glossary) formed by WGD. In practice, additional genomic rearrangements might disperse the orthologs across multiple chromosomes. Therefore, we rank the *Tetraodon* chromosomes according to how many orthologs of genes in the human *MATN3* neighborhood they contain, and we remove chromosomes containing fewer than six of these orthologs. Then, we define as co-orthologous chromosomes the smallest set of *Tetraodon* chromosomes that together contain at least 30% of the orthologs of human genes in the *MATN3* region. For *Tetraodon*, this procedure identifies chromosomes Tni10 and Tni14 as probable co-orthologs of the human *MATN3* region (Figure I). Applying the same method to zebrafish, we identify Dre17 and Dre20.

Next, we establish which of these *Tetraodon* chromosomes is the ortholog of which zebrafish chromosome. Because the WGD occurred before these teleosts diverged, and because the most frequent fate for the descendants of an ancestral locus is the rapid loss of one gene copy after WGD [8], we should observe a large number of gene losses that are common to both fish species, in addition to a smaller number of ancestral loci where both copies have been retained and, much more rarely, RGL. The number of ancestral loci where the surviving fish genes are located on two orthologous chromosomes should, therefore, greatly exceed the number of ancestral loci where they map to two paralogous chromosomes. We constructed an Oxford grid containing the number of orthologous genes located on each pair of chromosomes between *Tetraodon* and zebrafish (Appendix BTable S2 in the supplementary material online). This grid shows that the orthologs of 210 genes on *Tetraodon* chromosome Tni14 are located on zebrafish chromosome Dre20, and there are 90 orthologs for the pair Tni10 and Dre17. By contrast, for the alternative possible pairing, there are only 17 orthologs on Tni10 and Dre20, and 22 on Tni14 and Dre17. This analysis clearly identifies Tni14 and Dre20, and Tni10 and Dre17, as the two orthologous pairs among the *Tetraodon* and zebrafish chromosomes that are co-orthologous to the human *MATN3* region (Chi-square test, $P < 0.05$).

It is now possible to complete the tracking of the *MATN3* region. For the orthologs of human *MATN3*, the *Tetraodon* gene is located on Tni10, whereas the zebrafish gene is on Dre20. Because Tni10 and Dre20 are paralogous chromosomes, we infer that *MATN3* was still duplicated at the time that the *Tetraodon* and zebrafish lineages separated, and these duplicates subsequently underwent reciprocal loss.

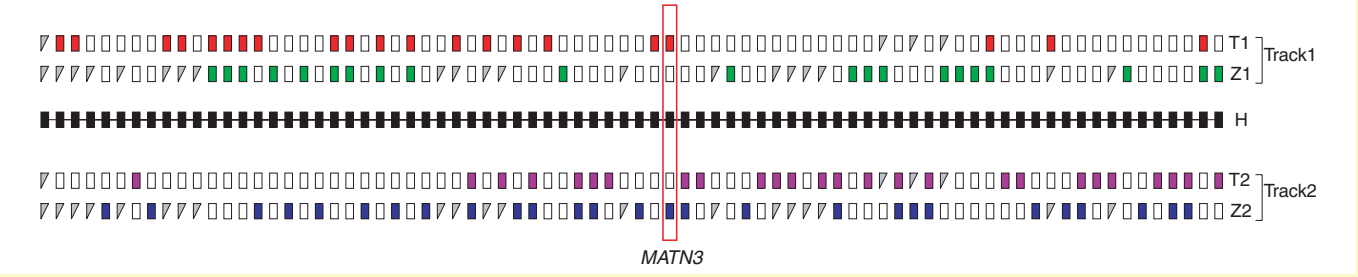

Figure I. Example of RGL between *Tetraodon* and zebrafish. At the *MATN3* locus, the duplicates have undergone RGL between *Tetraodon* and zebrafish. Human genes (H) are represented by black boxes on the central horizontal track, the line showing the immediate proximity of these genes on their chromosome. The two corresponding regions in each fish species – *Tetraodon* (T) and zebrafish (Z) – are represented by the two tracks at the top and bottom. The fish genes on each track are not necessarily linked. Each column corresponds to a set of homologous genes descended from one ancestral locus. Colored rectangles represent genes belonging to the fish chromosomes that we have annotated as homologous to this outgroup window. *Tetraodon* chromosome Tni10 is shown in red; and Tni14, in purple. Zebrafish (*Danio rerio*) chromosome Dre17 is shown in green; and Dre20, in blue. Gray triangles mark genes on other fish chromosomes, and white rectangles mark genes that have been lost. Genes located on fish chromosomes other than Tni10, Tni14, Dre17 and Dre20 cannot be assigned to either track and, therefore, are represented by gray triangles on both tracks. At the *MATN3* locus (red box), the duplicates have undergone RGL in *Tetraodon* and zebrafish.

any over-represented terms, differing from observations in yeast [8], but this finding is perhaps not surprising given the small size of the data set. We also checked whether expression breadth differed between RGL and non-RGL loci, using expressed-sequence tag (EST) data from humans and zebrafish, but found no significant difference (data not shown).

Testing the accuracy of RGL locus identification

To test the reliability of our definition of homologous regions between teleost species and outgroup, we bootstrapped the composition of the 80-gene window around the central locus under study (as defined in Box 1) 100 times. This means that, for each window, we sampled gene composition with replacement and assessed the status (i.e. RGL, non-RGL or not determined) of the central locus. This analysis confirmed the status for each locus in at least 97% of the replicates. We also carried out the same analysis using a more stringent definition of orthology between fish chromosomes and found that the estimated proportion of RGL among ancestral loci with identifiable single-copy descendants in both teleosts remained in the range 6.8–7.7% (Table 1, Orthology of fish chromosomes rows).

One potential problem with our analysis is that poor annotation of a fish genome sequence might result in some double-copy genes being misclassified as single-copy genes (and the locus might be counted either as an RGL or a non-RGL locus). To assess whether this is a considerable problem, we used BLAT (BLAST-like alignment tool) [22] to align each putative single-copy fish protein with genomic DNA of both fish, using various thresholds (Table 1, Annotation check rows). Although some of the identified RGL and non-RGL loci are removed by these filtering steps, most RGL loci are retained (80%), and the relative proportion of ancestral loci classified as RGL is unaltered. We might have missed some genes that map to gaps in the genome assembly, but the quality of the assembly is high, at least for *Tetraodon* [1]. Nonetheless, we do not think this

Table 1. Number of RGL loci in teleosts, determined using various criteria[a]

		Loci in outgroup	RGL loci	Non-RGL loci	Not determined	Proportion of RGL loci (%)[b]
Outgroup	Human	2124	44	566	1514	7.2
	Chicken	2259	42	642	1575	6.1
	Merged	2772	73	841	1858	8.0
Orthology of fish chromosomes[c]	Chi-square test ($P < 0.001$)[d]	2772	69	831	1872	7.7
	Chi-square test ($P < 10^{-5}$)[d]	2772	58	797	1917	6.8
	One-to-one orthologs[e]	2772	40	516	2216	7.2
Annotation check[c]	Match compatible with phylogeny[f]	NA	60	668	NA	8.2
	Match in appropriate locations[g]	NA	61	809	NA	7.0

[a]Abbreviation: NA, not applicable.
[b]The ratio defined by the following equation:(number of RGL loci × 100) ÷ (number of RGL loci + number of non-RGL loci).
[c]Analysis carried out on merged (chicken and human) outgroup data sets.
[d]Restriction to pairs of chromosomes for which the Chi-square test *P* value is <0.001 or $<10^{-5}$ (Box 1). The threshold 10^{-5} corresponds to a Bonferroni correction for multiple testing.
[e]Restriction to the 15 pairs of chromosomes with a clear one-to-one relationship between *Tetraodon* and zebrafish (light gray cells in Table S2 in the supplementary material online).
[f]Loci were discarded if we found, using BLAT [22], a match in any location with sequence similarity compatible with expectations given the phylogeny. The divergence of *Tetraodon* and zebrafish is more recent than the WGD, so orthologs (in different species) should be more similar than paralogs (in the same species). In the case of RGL loci, we measured the similarity between the previously annotated teleost proteins (encoded by paralogs in that case) and used this criterion to identify possible missing copies. In the case of non-RGL loci, we estimated the average similarity between paralogs using 34 families in which all four possible teleost genes had been retained after the WGD. We used the confidence interval of this value to look for matches for which the similarity was compatible with these observations.
[g]Because unannotated genomic regions could contain fast-evolving gene copies, we carried out a low-stringency search, and we discarded any locus for which a match with more than 40% identity over more than 30% of the query length was found in the appropriate syntenic region (defined as described for the matrilin 3 locus; Box 1). Loci with any match mapped on the appropriate chromosomes were discarded.

could influence the ratio of ancestral loci classified as RGL to those classified as non-RGL, and we are confident that we have eliminated the possibility that most of the apparent occurrences of RGL might be annotation artifacts. We conclude that the descendants of a non-negligible proportion of ancestral loci have undergone RGL since the divergence between *Tetraodon* and zebrafish. This proportion is estimated by various methods to be between 6.1% and 8.2% (Table 1).

Conclusions

Several examples of RGL in medaka (*Oryzias latipes*) and zebrafish have been reported previously [19], but this study did not control for unannotated genes or gaps in the genome sequence, which are crucial factors, because the medaka genome sequence is incomplete at present. Our study is the first large-scale analysis of RGL in fish. Although it has been proposed that there is a temporal correlation between the WGD and the radiation of teleost species [2,3,5,7], this relationship is controversial, because it disappears when extinct lineages are taken into account [23] and because the time frames of both the WGD and the teleost radiation are uncertain. Our results show that duplicated genes formed by the WGD were still being lost at the time of the last common ancestor of *Tetraodon* and zebrafish. Because the RGL loci show no evidence of functional bias, we are not suggesting that they mainly consist of any particular types of gene or that they have been subject to any unusual form of natural selection. Instead, we suggest that they are simply genes that did not need to be retained in duplicate and that happened to undergo reciprocal loss in the two teleost lineages studied here.

Assuming that the ancestral loci that we were able to study here (owing to their conserved syntenic relationship to outgroups and among teleosts) are representative of the whole genome, we estimate that the descendants of ~1700 ancestral loci (i.e. ~7% of ~25 000 loci in the pre-WGD fish ancestor) have undergone RGL in *Tetraodon* and zebrafish. Because RGL at only 20–30 locus pairs encoding essential genes is sufficient to result in reproductive isolation by a Dobzhansky–Muller process [9,11], this estimate implies that RGL at duplicated loci is probably a contributing factor to all speciation events that occurred between the teleost WGD and some time after the *Tetraodon* and zebrafish divergence. Ongoing RGL would have continued to split lineages for tens of millions of years after the WGD, until such time as the rate of gene loss slowed [8] to the point at which the genomes of mating individuals were unlikely to differ by RGL at more than 20–30 ancestral loci.

RGL is a particular form of divergent resolution after gene duplication, in which only paralogs are retained. Another possible outcome from an ancestral locus is the retention of two copies in one species but only one copy in the other species. Examples of such two-to-one relationships have been reported between *Tetraodon* and zebrafish [24] and might have a similar impact on speciation if the two copies that were retained undergo subfunctionalization. Theoretical developments [25,26] have extended the link between RGL and speciation to the case of subfunctionalization, in which both copies survive after WGD and each keeps a subset of the functions of the ancestral gene. Similar to RGL, the differential partition of gene expression between duplicated gene copies in different populations might promote reproductive isolation. One example of differential subfunctionalization in stickleback (*Gasterosteus aculeatus*) and medaka has been described [6,26], but the extent of this phenomenon has not yet been determined. Therefore, the potential for gene pairs formed by WGD to contribute later to speciation events extends beyond the loci implicated here.

Acknowledgements

This study was supported by the Irish Research Council for Science, Engineering and Technology and by Science Foundation Ireland. We thank Meg Woolfit and Brian Cusack for a critical reading of the manuscript, and all of the members of our laboratory for help with the Glossary. The zebrafish genome is being sequenced and annotated by the Wellcome Trust Sanger Institute. The whole project is funded by the Wellcome Trust, and all data can be accessed from the web page at http://www.sanger.ac.uk/Projects/D_rerio/

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.01.003.

References

1 Jaillon, O. *et al.* (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946–957
2 Christoffels, A. *et al.* (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* 21, 1146–1151
3 Vandepoele, K. *et al.* (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1638–1643
4 Brunet, F.G. *et al.* (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol. Biol. Evol.* 23, 1808–1816
5 Hoegg, S. *et al.* (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* 59, 190–203
6 Cresko, W.A. *et al.* (2003) Genome duplication, subfunction partitioning, and lineage divergence: *Sox9* in stickleback and zebrafish. *Dev. Dyn.* 228, 480–489
7 Amores, A. *et al.* (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714
8 Scannell, D.R. *et al.* (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440, 341–345
9 Lynch, M. and Force, A. (2000) The origin of interspecies genomic incompatibility via gene duplication. *Am. Nat.* 156, 590–605
10 Taylor, J.S. *et al.* (2001) Genome duplication, divergent resolution and speciation. *Trends Genet.* 17, 299–301
11 Werth, C.R. and Windham, M.D. (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* 137, 515–526
12 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
13 Hubbard, T. *et al.* (2005) Ensembl 2005. *Nucleic Acids Res.* 33, D447–D453
14 Benton, M.J. and Donoghue, P.C. (2007) Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 24, 26–53
15 Kumazawa, Y. *et al.* (1999) Mitochondrial molecular clocks and the origin of euteleostean biodiversity: familial radiation of perciforms may have predated the Cretaceous/Tertiary boundary. In *The Biology of Biodiversity* (Kato, M., ed.), pp. 35–52, Springer-Verlag
16 Inoue, J.G. *et al.* (2005) The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* 349, 227–235
17 Yamanoue, Y. *et al.* (2006) The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet. Syst.* 81, 29–39
18 Hillier, L.W. *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716
19 Naruse, K. *et al.* (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* 14, 820–828
20 Kellis, M. *et al.* (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624
21 Al-Shahrour, F. *et al.* (2004) FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics* 20, 578–580
22 Kent, W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Res.* 12, 656–664
23 Donoghue, P.J.C. and Purnell, M.A. (2005) Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* 20, 313–319
24 Woods, I.G. *et al.* (2005) The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.* 15, 1307–1314
25 Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473
26 Postlethwait, J. *et al.* (2004) Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 20, 481–490

0168-9525/$ – see front matter
doi:10.1016/j.tig.2007.01.003